



TARTU ÜLICOOL

Loodus- ja Tehnoloogiateaduskond

Ökoloogia ja Maateaduste Instituut

Geograafia Osakond

Bakalaureusetöö kartograafias ja geoinformaatikas

Ristproгноoside kasutatavus Otepää ja Karula taimkattetuvastuses

Eerik Absalon

Juhendaja: *MSc* Madli Linder

Kaitsmisele lubatud:

Juhendaja: *Linder*

Osakonna juhataja:

Tartu 2008

Sisukord

1. SISSEJUHATUS	3
2. UURIMISALADE KIRJELDUS	4
3. ANDMED.....	7
3.1. VÄLIVAATLUSANDMED	7
3.1.1. Välivaatluste kogumine.....	7
3.1.2. Õpetusvalimite koostamine.....	8
3.2. ANDMEKIHID JA NENDE ETTEVALMISTAMINE	9
3.2.1. Satelliidipildid	10
3.2.1.1. NDVI andmekihid.....	13
3.2.2. Põhikaardiandmed.....	14
3.2.2.1. Põhikaardi rasteriseerimine.....	14
3.2.2.2. Põhikaardi metsaliigid	14
3.2.2.3. Kaugus vooluveekogudest	14
3.2.3. Mullakaardi rasteriseerimine	15
3.2.4. Maapinna kõrgusmudel SRTM	15
3.2.5. Ortofotod ja asendustunnuste kihid	16
4. PIDEVSTUUDIUMI METOODIKA	16
4.1. RUUMIMUSTRIT KIRJELDAVAD TUNNUSED	17
4.2. TEHISÕPE	18
4.3. HINNANGUTE ARVUTAMINE	18
4.3.1. Hinnangud andmebaasi tabelisse.....	18
4.3.2. Hinnangud binaarfaili	19
5. TULEMUSED	19
5.1. TEHISÕPPEL VÄLJA VALITUD TUNNUSTEST	19
5.2. HINNANGUD TABELISSE	20
5.3. PROGNOOSIKAART	22
6. ARUTELU	26
7. KOKKUVÕTE.....	28
TÄNUAVALDUSED	28
SUMMARY	30
VIIDATUD KIRJANDUS	31
LISAD	36

1. Sissejuhatus

Tänapäeval kasutatakse looduses esinevate objektide äratundmiseks ja kirjeldamiseks teaduses järjest rohkem kaugseire abi. Kaugseire on objekti, ala või nähtuse kohta info hankimine, kasutades selleks salvestusvahendeid, mis pole olemata uuritava objekti, ala või nähtusega otseses kontaktis (Lillesand, Kiefer, 2000). Eriti kasulikud on kaugseireandmed siis, kui soovitakse teha suuremaid piirkondi hõlmavaid kaarte, kuna välitööd on aja- ja ressursinõudlikud ning kaugseire kasutamine välitööde asemel aitaks kulusid suuresti vähendada (Kilpeläinen, Tokola, 1999). Nagu mujal maailmas, on ka Eestis kaugseireandmeid kasutatud taimkatte ja maakatte kaardistamiseks ning metsa inventeerimiseks juba pikka aega (Aaviksoo, 1995; Aaviksoo *et al*, 1997; Aaviksoo *et al*, 2000; Kuusk *et al*, 2004; Matsalu...2007; Meiner, 1999; Nilson *et al*, 1999; Palo, 1999). Sageli kombineeritakse kaugseireandmeid samal eesmärgil muude andmestikega, nt kõrgusmudelid (Kimes *et al*, 1996; Mather, 1999), mis annab sageli paremaid tulemusi.

Üks võimalus taimkattetüüpide kaardistamiseks, mida on kasutatud ka käesolevas bakalaureusetöös, on väliandmete ning kaugseire- ja kaardiandmete kui seletavate tunnuste andmekihtide kombineerimine hinnangulisel kaardistamisel näidistel põhineva järeldamise ja pidevõppe metoodika abil.

Bakalaureusetöö eesmärgid olid:

- 1) teha endale selgeks näidistele tugineva tehisõppesüsteemi *Pidevstuumium* ülesehitus, selles kasutatavate andmekihtide ettevalmistamise ning hinnanguliste taimkattekaartide loomise protseduurid;
- 2) uurida, kui hästi on ühe uurimisala vaatlusandmete põhjal võimalik prognoosida teise, eemalasuva ala taimkatet (Otepää piirkonna põhjal Karulat ning vastupidi).

2. Uurimisalade kirjeldus

Uurimus viidi läbi kahes Lõuna-Eesti piirkonnas. Esimene neist hõlmab peamiselt Otepää kõrgustiku ja looduspargi lõuna- ja lääneosa, ulatudes vähesel määral ka naabermaastikurajooni Ugandi lavamaale (Eesti põhikaardi lehed 5423, 5424, 5433, 5443; uurimisala nimetus edaspidi: Otepää). Teine uuritav piirkond katab täielikult Karula Rahvuspargi ala ning suurema osa Karula kõrgustikust (Eesti põhikaardi lehed 4494, 4495, 5404, 5405; edaspidi: Karula) (joonis 1). Otepää kõrgustik otsustati uurimusse kaasata, kuna Otepää Looduspargi taimkatet on käesolevas uurimuses kasutatava metoodika abil juba varem kaardistatud (Remm *et al.*, 2007). Võrreldavaks piirkonnaks sobis aga hästi Karula kõrgustik, mis on geneetiliselt (Granö, 1922, *cit* Raukas, Rõuk, 1995) ning maastikuliselt Otepää kõrgustikuga seotud. Ka paigastike levimuses võib leida ühiseid jooni. Piirangud uurimisalale seadis võrreldavate andmekihtide olemasolu mõlemas piirkonnas, s.t., kuna sama aasta (2005) Karula ortofotod olid olemas vaid teatud Otepää ala kohta, siis seda ala uuritigi.

Liigestatud pinnamoe tõttu on Karula ja Otepää kõrgustiku taimkate väga mosaiikne. Liigestatud pinnavormiga aladel nagu Otepää kõrgustik ja Karula kõrgustik esineb palju deluviaalseteid mis on tekkinud nõlvadelt ärauhutud materjali koondumisel nõgudesse, kus nad siis hiljem on soostunud – sellest ka suur soiste või soostuvate alade rohkus (Raukas, 1995). Pinnamoelt on Karula ja Otepää künklikud kõrgustikud.

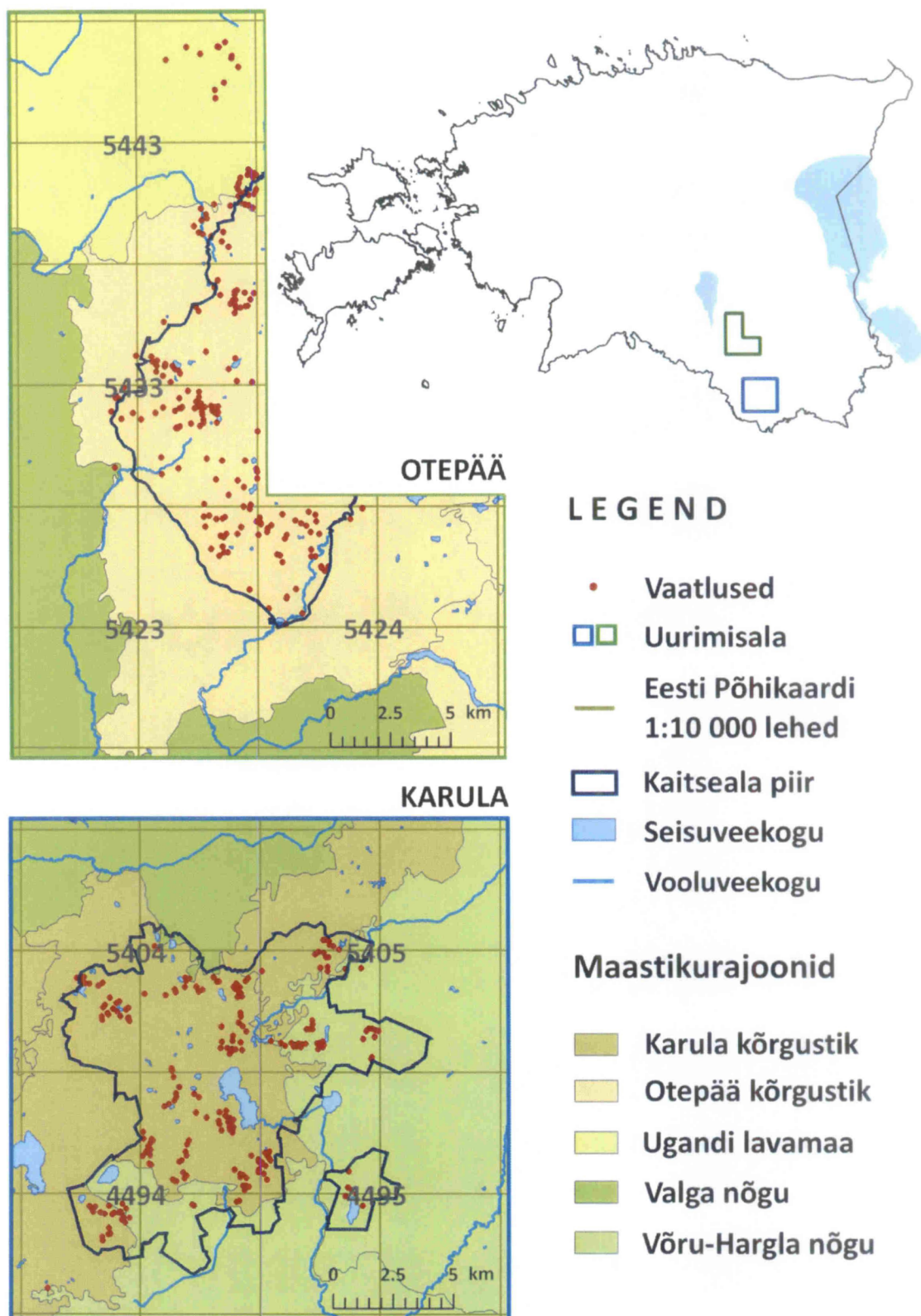
Otepää kõrgustiku pindala on 1247 km² (Arold, 2005) ja keskmine kõrgus 127 meetrit (Hang, 1979). Kõrgeim tipp on kõrgustiku lääneosas asuv Kuutse mägi kõrgusega 217 meetrit. Geneesilt on Otepää kuhjeline saarkõrgustik. Peaaegu poolt kõrgustikku katavad moreenkünkad ja moreenkattega mõhnad, kvaternaarisetete paksus on kohati kuni 200 meetrit, enamasti umbes 50 meetrit. Kõige paksem on see mattunud ürgorgudes (Kooraste, Urvaste, Voki-Truuta), mis esinevad peamiselt kõrgustiku äärealadel ning mis on väga järvederohked (Raukas, 1995).

Karula kõrgustiku pindala on 275 km². Ta on madalam kui Otepää kõrgustik, tipud on enamasti 110-120 meetri kõrgused ja kõige kõrgem on Rebasejärve Tornimägi (137 m) (Arold, 2005). Geneesilt on Karula kõrgustik mandrijää serva esine kuhjekõrgustik. Seda kutsutakse ka Karula kuplistikuks. Kvaternaarisetete paksus on 20-60 meetri kanti. Karula kõrgustikul on Kagu/Lõuna-Eestis enim üle 1 ha suurusi järvi.

Mõlemad, Otepää ja Karula kõrgustik, jäävad mullastiku poolest Otepää-Karula allvaldkonda ehk erodeeritud kamar-leetmuldade (moreenküngastike) valdkonda. Otepääl esineb ka leostunud ja erodeeritud kamar-karbonaatmuldi. Lähtekivimiks on olnud punakaspruun põhimoreen mis koosneb saviliivadest ja liivsavidest (Laasimer, 1965).

Tänapäeval kasvavad Otepääl liigivaesed kuusikud ning väikeselehised metsad, Karulas aga paluka- ja samblikumännikud ning väikeselehised metsad (Laasimer, 1965). Aroldi (2005) järgi levivad Karulas kuusikud ja nendega suktessioonireas vahelduvad lehtpuuenamusega segametsad. Üldiselt võib Karulast leida kõiki metsatüüpe sürjametsadest siirdesoometsadeni. Otepääl kasvavad aga liigirikkad sürjametsad, laanemetsad, palumetsad, sambliku- ja kanarbikumännikud ning mitmed erinevad kuusikud.

Otepää kõrgustiku lääneosas paikneb Otepää Looduspark mis on suurim maastikukaitseala Eestis – 224,3 km² (Otepää..., 2006). Ka Karula kõrgustikul asub kaitseala – Karula Rahvuspark, mis on Eesti väikseim rahvuspark pindalaga 110 km² (Arold, 2005).



Joonis 1. Uurimisalade paiknemine.

3. Andmed

3.1. Välivaatlusandmed

3.1.1. Välivaatluste kogumine

Otepää uurimisalalt kasutati TÜ geograafia osakonnas olevas *MS Accessi* vormingus taimkattenäidiste andmebaasis olevaid, aastatel 2002 ning 2004-2007 tehtud vaatlusi, millest uuritavatel kaardilehtedel paikneb 206 vaatlust. Karula andmed (243 vaatlust) koguti andmebaasi 2007. aasta 9. juulist kuni 19. oktoobrini ning neist 127 kogumisel osales ka autor. Lisaks on bakalaureusetöös kasutatavatesse valimitesse sattunud andmeid on kogunud Madli Linder, Eerik Absalon, Kalle Remm, Tõnis Relvik, Anneli Palo, Jaanus Remm, Elle Meier, Kristiina Sulbi, Õie Nikkel ja Jaanus Paal.

Algselt määrati ortofotode põhjal punktid, mis olid huvipakkuvad ja homogeensed, seejärel sisestati nende punktide koordinaadid käsiGPS-i. Kasutades GPS-i, kompassi ja kaartide abi, tehti kindlaks vaatluspunkti asukoht ning kirjelduse kehtivuse piirid looduses. Kohapeal märgiti üles kuupäev, punkti number, vaatleja(d), kasvukohatüüp, puittaimede katvus, 10-palline puistu valem maksimaalselt kolmes puurindes, põõsarinde koosseis ja alustaimestikku moodustavad liigid. Enamikus kohtades tehti ka kooslustest fotod. Kasvukohatüübid määrati Paali (1997) kasvukohatüüpide klassifikatsiooni järgi.

Puistu koosseisu ja katvust hinnati ülaltvaate järgi, kuna kaugseirepiltidel on ülaltvaade ning see võimaldab lihtsamini siduda välitööde andmeid kaugseire omadega. Katvust ülaltvaate järgi on kasutatud kõigis *Pidevstuudiumiga* seotud taimkatet puudutavates töödes (nt Linder, 2006; Remm, 2004, 2005; Remm *et al*, 2007; Oviir, 2007).

Näidisalasid valiti enamasti nelja kriteeriumi järgi:

- 1) ruumiline homogeensus (näidisala kirjeldus peaks kehtima vähemalt 20 m raadiuses);
- 2) ajaline stabiilsus (välditi hiljuti inimtegevuse tõttu või muul põhjusel muutunud kohti);
- 3) tüpoloogiline esinduslikkus (elistati taimkatteüksuste tüüpilisemaid esinemisvorme);
- 4) esindatus näidistebaasis (elistati taimkatteüksusi, mille esinemisvormid olid näidistebaasis senini vähe esindatud).

Karulas raskendas tervelt alalt andmekogumist liikumiskeeld Pautsjärve loodusreservaadis ning ajutised liikumiskeelud Palotaguse, Äestaja soo, Oekonnu, Rebase, Vana-Palo ja Põrgujärve sihtkaitsevööndites (Keskkonnaministeerium, 2008). Mõnest nimetatud piirkonnast koguti andmeid pärast liikumiskeelu lõppu.

Vaatlusala tunnusnumber ja piirid digitaliseeriti *MicroStationi* faili ning kontrolliti kameraalselt, et ala kirjeldus oleks vastavuses ortofoto ja põhikaardiga.

3.1.2. Õpetusvalimite koostamine

Kokku oli kasutatavate vaatluste andmebaasis 449 vaatlust, mille põhjal moodustati valimid. Valimid koostati kihilise juhuvalikuga *MS Excelis*. Kihiline valik (*stratified sampling*) on valikumeetod, mille puhul andmed (vaatlused) klassifitseeritakse enne valimi moodustamist olulis(t)e tunnus(t)e järgi kihtidesse, mille piires kasutatakse seejärel enamasti juhuvalikut. Meetod tagab uurimuse jaoks oluliste andmekihtide esindatuse (Remm, 2008) ning annab ülevaate olemasolevatest andmetest (Bartolucci *et al*, 2004). Kihilise juhuvalimi jaoks grupeeriti vaatlused kolme üldisesse klassi (tabel 1).

Tabel 1. Vaatluste jaotumine klassidesse kihilisel juhuvalikul. Värvilisega on märgitud kasutatavate vaatluste arv valimite moodustamiseks.

Klassi nimi	Vaatluste arv	
	Karula	Otepää
Niidud ja sood	27	42
Haritav ja kultuurmaa	27	30
Metsad	189	134
Kokku	243	206

Mõlema uurimisala kõikidest andmetest moodustati kaks teineteist välistavat juhuvalimit. Valimite suuruse määrasid ära eelnevalt loodud üldisemate vähima vaatluste arvuga klasside mahud, s.t. 27 vaatlusega kihi puhul sai moodustada kaks 13-vaatluselist valimit ning 134 vaatlusega kihi puhul kaks 67 vaatlusega valimit (igas valimis seega kokku 93 vaatlust). Lisaks moodustati kaks teineteist välistavat Otepää-Karula 93 vaatlusega segavalimit. Kõigist vaatlustest 44 ei sattunud ühtegi juhuvalimisse. Lisaks neile moodustas *Pidevstuudium*

tehisõppe käigus üksikud juhuvalimid mahuga ~93 punkti kõigist Karula 243 vaatlusest, kõigist Otepää 206 vaatlustest ja kogu andmestikust, s.o. 449 vaatluse hulgast. Ühesuuruste kihiliste valimite kasutamine peaks eeldatavalt välistama valimi suurusest tulenevad erinevused erinevate valimitega saadud hinnagute täpsuses.

Kõik vaatlused jagati 11 funktsioontunnuse (hinnatavasse) klassi. Vaatluste jaotus valimites on toodud tabelis 2.

Tabel 2. Vaatluste arv funktsioontunnuse klassides valimite kaupa. *kar1*, *kar2* – 93-vaatluselised üksteist välistavad Karula juhuvalimid, *ote1*, *ote2* – 93-vaatluselised üksteist välistavad Otepää juhuvalimid, *ote&kar1*, *ote&kar2* – 93-vaatluselised üksteist välistavad Otepää ja Karula segavalimid.

klassi kood	klassi nimetus	kar1	kar2	ote1	ote2	ote& kar1	ote& kar2
1	looduslikud kuivad ja niisked niidud	3	3	6	3	4	7
2	looduslikud märjad ja soostunud niidud	3	2	1	4	1	4
3	sood	7	8	6	6	8	2
4	söödid	3	1	7	8	4	6
5	haritav ja kultuurmaa	10	12	6	5	9	7
6	nõmme-, laane- ja palumetsad	28	28	12	13	27	23
7	sürja- ja salumetsad	13	5	36	31	18	14
8	soostuvad ja rabastuvad metsad	5	8	4	2	3	5
9	pajustikud	1	3	1	3	1	
10	soometsad	17	16	11	14	16	19
11	kõdusoometsad	3	7	3	4	2	6

3.2. Andmekihid ja nende ettevalmistamine

Pidevstuudiumi tehisõppes ja prognoosis kasutatavad andmekihid peavad vastama kindlatele nõuetele: kõik kihid peavad olema *Idrisi32 rst*-formaadis, iga andmekiht peab asuma omaette, andmekihtide andmebaasis kirjas oleva nimega kataloogis, kõik rasterfailid peavad olema Eesti andmete puhul põhikaardi projektsioonis ja ristkoordinaatide süsteemis (LAMBERT-EST97) ja iga fail esindama ühte 10 × 10 km kaardilehte, faili nimeks peab olema kaardilehe neljakohaline number laiendiga *rst* (näiteks *4494.rst*) jne (Remm, Linder, 2008).

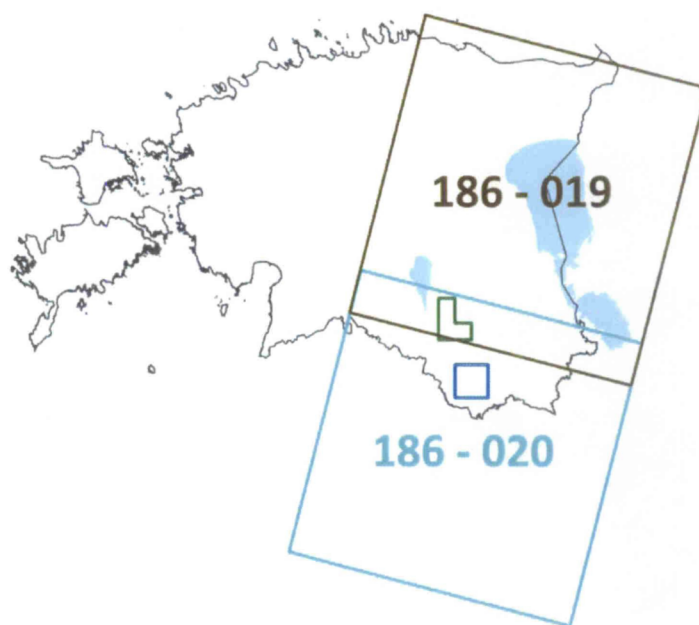
Järgnevalt on kirjeldatud uurimuses kasutatud andmekihtide ettevalmistamist *Pidevstuudiumile* vajalikule kujule. Andmekihid valmistati ette enamjaolt uurimuses kasutatud kaardilehtede kohta, satelliidipiltide puhul valmistati ette kõik kaardilehed, mida hõlmasid Landsat 5 kasutatud kaadrid (joonis 2).

3.2.1. Satelliidipildid

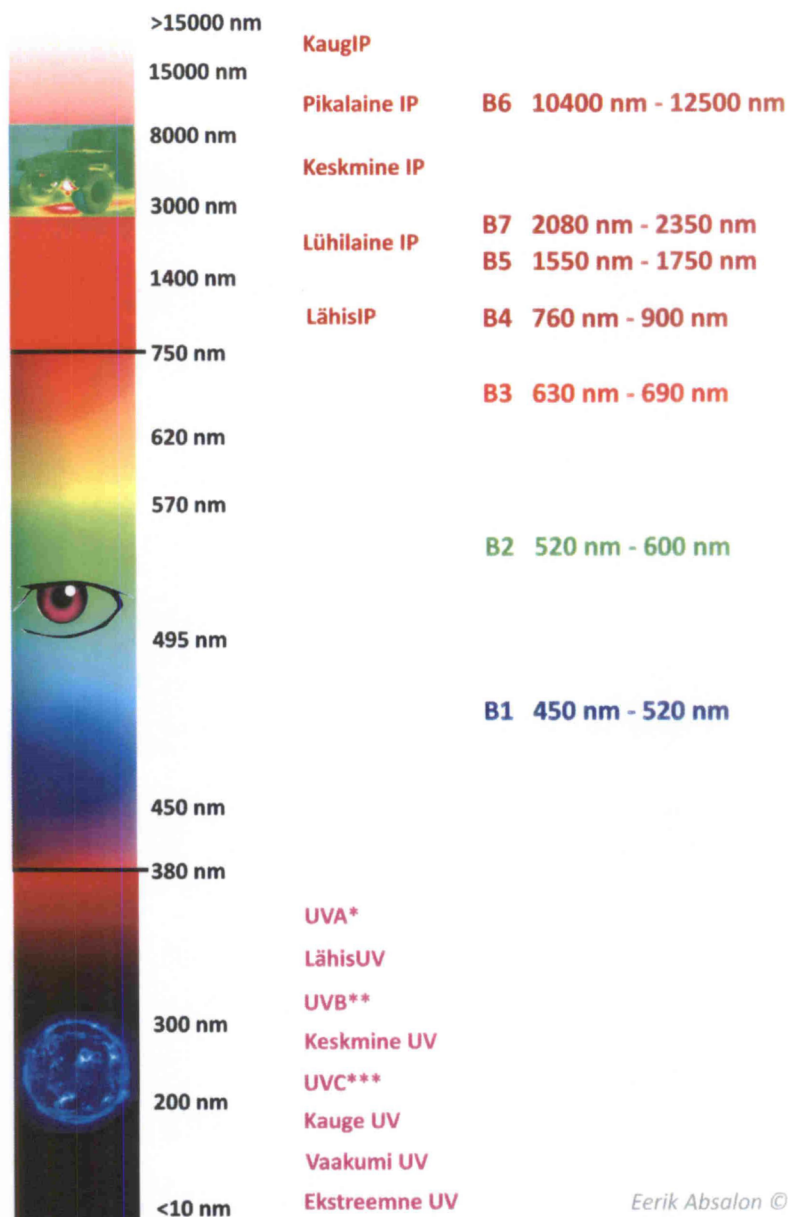
Satelliidipiltidest kasutati Landsat 5 TM-i 21. mai ning 9. augusti 2007 kaadreid 186-019 ning 186-020 (joonis 2). Kaadrid katavad Ida- ja Lõuna-Eesti, väikese osa Venemaast ning Lätist, seda seitsmes erinevas spektraalkanalisis (joonis 3).

Landsati programm on seni pikaajaliseim ettevõtmine omandamiseks pilte Maast kosmosest. Esimene satelliit lasti üles 1972. aastal ja viimane (Landsat 7) aastal 1999 (LANDSAT General...). Landsat 5 on Landsat 4 koopiat ja ta ehitati sellele tagavaraks. Mõlema pardal on TM (*Thematic Mapper*) ja MSS (*Multi Spectral Scanner*) (Lillesand, Kiefer, 2000; Proosa, 2007).

Landsat 5 lennutati orbiidile 1. märtsil 1984. Tal kulub 16 päeva, et skanneerida terve Maa. 2005 tekkisid aga satelliidil probleemid päikesepatareidega, mis ei lasknud akudel laadida ja tema operatsioon peatati, seetõttu jäi seisma ka piltide tegemine. Jaanuaris 2006 aga saadi satelliit töökorda ja ta võis oma tööd jätkata (Landsat 5..., 2008).



Joonis 2. Landsat 5 TM kaadrite 186-019 ja 186-020 paiknemine Eesti alal ning uurimisalade suhtes. Otepää uurimisala – roheline kujund, Karula uurimisala – sinine ruut.



* Ultraviolett A, pikalaineline, ka tuntud kui „black light“

** Ultraviolett B, keskmiselaineline

*** Ultraviolett C, lühilaineline, antibakteriaalne

Joonis 3. Landsat 5 Thematic Mapper'i (TM) seitsme kanali (B1...B7) lainepikkused elektromagneetilises spektris. Kanalite resolutsioon on 30×30 m, B6-l 120×120 m.

Satelliidipildid on *.tif formaadis. TIFF on ainus üldkasutatav formaat, mis võimaldab kompressiooni, mosaiikimist ja laiendusi sidumaks rasterpilti geograafiliste andmetega (GeoTIFF...). Satelliidipiltide ettevalmistamiseks Pidevstuumile vajalikule kujule kasutati programmi Idrisi32. Failid imporditi Idrisisse mooduli GEOTIFF abil. Samade kuupäevadega kahe kaadri pildid liideti kanalite kaupa. Liidetud pildid said vastavalt oma kanali nime – B1...B7 (B – band, kanal). Seejärel transformeeriti satelliidipildid algsest projektsioonist (UTM) Eesti põhikaardi projektsiooni. Projektsiooni teisendamisel kasutati kindelpunktide (ground control points - GCP) (Idrisi32...) asukohtade ning koordinaatide määramisel

abimaterjalina 2005. aasta ortofotosid ning Eesti põhikaarti. Kindelpunktideks ei sobi kumerad objektid, kaared, vaid pigem teravamad nurgad nagu teede ristumiskohad – need on inimtekkelised ja seetõttu piltidel paremini tuvastatavad (Lyons, 2002; Zhen *et al.* 2004). Kindelpunktid peaksid olema paigutatud ühtlaselt üle pildi/kaardi/foto, et vastavasse projektsiooni vorpides tervet objekti ühtlaselt muudetaks ning et vead töötlemise käigus võimalikult väikesed tuleks. Minimaalseks kindelpunktide arvuks loetakse lineaarse teisenduse puhul neli punkti. Kindelpunktide leidmiseks kasutati Landsat 5 seitsmenda ehk keskinfrapunase kanali (2.08-2.35 μm) liitkaadrit, sest sellel eristusid teed muust maastikust kõige paremini. Mõlema pildi (mai ja augusti) jaoks leiti 20 sidumispunkti, mille hulgast suurema veaga punktide järjestikuse eemaldamisel jäi järgi kaheksa suhteliselt ühtlaselt üle liitkaadri paiknevat punkti. Augustikuise satelliidipildi puhul saavutati RMSE 3.25 ja maikuu puhul 3.59. Pärast teisendust kontrolliti täpsust veel ortofoto alusel.

Pärast seda, kui satelliidipildid olid sobivaks elastitud, lõigati need mõõtmetelt (10 × 10 km) ja koordinaatidelt Eesti põhikaardi lehtedele vastavateks tükkideks. Kuna *Pidevstuudium* kasutab andmekihte, mis paiknevad kindlaksmääratud nimega kaustades, siis tuli nad kõik vastavalt oma kanali nimega (B1...B7) kausta liigutada ja ümber nimetada, et faililaiendi ette jääks alles ainult põhikaardilehe neljakohaline number (nt 4494). Kuna kahe kuupäeva kohta tuli kaardilehe faile kokku üle 5000, kasutati nende ümberpaigutamiseks ja -nimetamiseks järgmist koodijuppi:

```
cmd /v /c "for %a in (*.*) do set abc=%~na && cd !abc:~24! && move ..\%a  
!abc:~0,4!%~xa && cd .."
```

Protsess, mis hinnanguliselt pideva töö jooksul oleks aega võtnud üle 15 tunni, võttis koodil aega vaid paarkümmend sekundit. Kirjeldatud kood eeldab, et failinimi on 26 tähemärki pikk ja viimased kaks neist peavad viitama kanalile (B1...B7). Kaustad, kuhu nad liigutatakse tuleb eelnevalt ette valmistada, selle töö puhul olid nendeks kaustad B1 kuni B7. Esimesed neli tähemärki aga peavad viitama põhikaardilehele ja need jäetakse ümbernimetamise käigus alles.

Näide kirjeldatud koodiga muudetavast/liigutatavast failipaarist:

4494RESAMPLE-09-08-2007-B1.RDC

4494RESAMPLE-09-08-2007-B1.rst

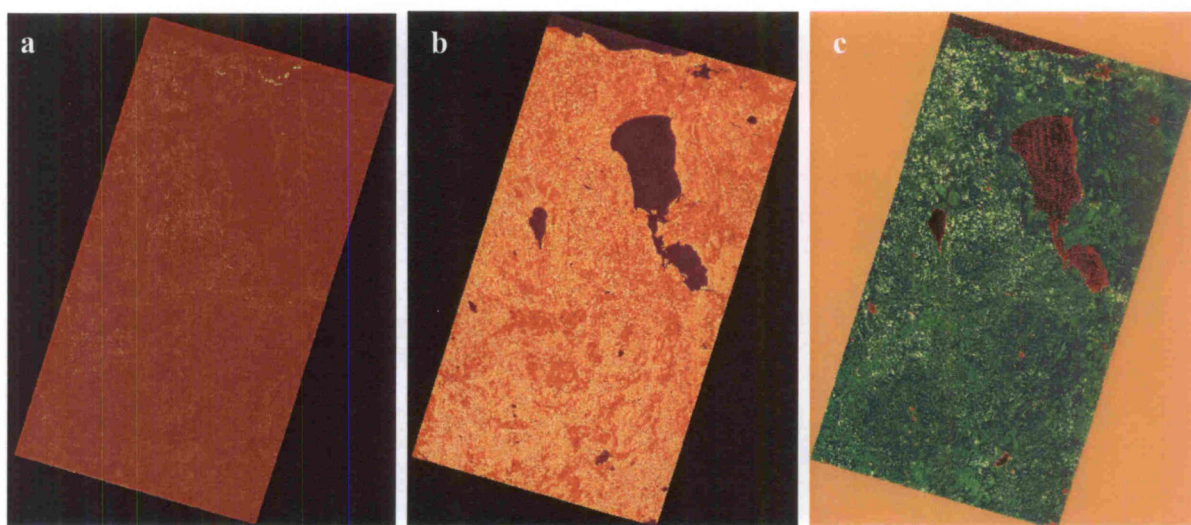
3.2.1.1. NDVI andmekihid

NDVI (*Normalized Difference Vegetation Index*) on kaugseire pildiandmetest tuletatud indeks, mis kirjeldab taimkatte elujõulisust, fotosünteesi efektiivsust, biomassi hulka. Selleks kasutatakse kanali B3 (punane) ja kanali B4 (lähisinfrapunane; valemis: *LähisIP*) kombinatsiooni (joonis 4), kuna nendes kiirgusvahemikes tuleb kõige paremini esile mulla ja taimkatte kontrastsus (Idrisi32...; Lillesand, Kiefer, 2000). NDVI arvutamiseks kasutatakse valemit:

$$\text{NDVI} = \frac{B4 - B3}{B4 + B3} \quad \text{ehk} \quad \text{NDVI} = \frac{\text{LähisIP} - \text{Punane}}{\text{LähisIP} + \text{Punane}} \quad (1)$$

Kuna see annab pikslitele ka negatiivsed ja väga väikesed (<0) väärtused, *Pidevstuudiumi* andmekihid aga peavad olema baidiformaadis, siis kasutati lõpliku kihi saamiseks *Idrisi32* pildikalkulaatorit ning teisendamist baidiformaati. Lõplik valem oli järgmine:

$$\text{NDVI} = \frac{\text{LähisIP} - \text{Punane}}{\text{LähisIP} + \text{Punane}} * 100 + 100 \quad (2)$$



Joonis 4. Lansat 5 TM kaadrite 186-019 ja 186-020 liitpilt Ida-Eestist; a) kanal B3 (punane), b) kanal B4 (LähisIP) ja c) nendest arvutatud ning töödeldud NDVI andmekiht.

3.2.2. Põhikaardiandmed

3.2.2.1. Põhikaardi rasteriseerimine

Kasutatavad MicroStationi vormingus põhikaardilehed on vaja rasteriseerida, sest töös kasutatav programm *Pidevstuudium* oskab kasutada ainult rasterandmeid. Seda tehti vastavalt *Pidevstuudiumi* juhendile (Remm, Linder, 2008). Programmi *Idrisi32* kasutades teisendati 5 m resolutsiooniga põhikaardi põhialade andmekiht 10 m küljepikkuseks, eesmärgiga kasutada seda interpoleerimiskihina prognoosikaardi arvutamisel. Interpoleerimiseraldised ehk antud juhul põhikaardi põhialad piiravad kaartide genereerimisel tulemuste interpoleerimist arvutatud pikslite vahele.

3.2.2.2. Põhikaardi metsaliigid

Põhikaardi metsaliikide lugemiseks ja nendest andmekihi moodustamiseks kasutati vastavat õpetust *PIDEVSTUUDIUMi* juhendis (Remm, Linder, 2008).

3.2.2.3. Kaugus vooluveekogudest

Kraavide mõjuulatuse leidmine aitab eeldatavalt paremini ära tunda kõdusoometsi.

MicroStationist eksporditi põhikaardi *.oll failide vooluveekogusid sisaldavad kihid (Eesti põhikaardi..., 2002) dxf-formaati. Seejärel imporditi failid mooduli DXFIDRISI abil Idrisisse. Imporditud vektorfail konverteeriti rasterfailiks. Kraavidest kauguste leidmiseks kasutati Idrisi kaugusoperaatorit. Seejärel eemaldati baidiformaadi säilitamiseks ümberklassifitseerimise teel kõik 254st suuremad väärtused. Kaugus 254 meetrit ning enam on vahemaa, kus vooluveekogu mõju praktiliselt puudub.

3.2.3. Mullakaardi rasteriseerimine

1: 10 000 digitaalse mullakaardi *.rst failid olid vajalike kaardilehtede kohta küll andmekihtide hoidlas juba olemas, kuid vigade tõttu rasteriseeriti need uuesti.

Vead tulenesid järgmistest asjaoludest:

- 1) *MicroStationi* vormingus *.mld failidel puudus info mulla tunnusnumbri kohta faili päises, kust teisenduseks kasutatav *ArcView* seda otsib;
- 2) *ArcView* *.shp vormingus kaardilehtedel oli ülearuseid või puudulikke polügone – kaardilehtedel 44941 ja 44944 olid vastavalt küljes mullapolügonid lehtedelt 44951 ja 44954, millelt need omakorda puudusid;
- 3) programmi *Mullarasteriseerimine* kasutatavast andmebaasist olid mitmed mullakaardil esinevad MSLINK-koodid ning neile vastavad mullaliigid puudu;
- 4) üks mullaliik oli valesti klassifitseeritud.

Vigade vältimiseks kasutati mullakaardi rasteriseerimisel *.mld failide asemel *.shp-sid. Rasteriseerimine toimus vastavalt *Pidevstuudiumi* juhendile (Remm, Linder, 2008). Kasutati programme *ArcView* ja *Mullarasteriseerimine*. Üldiselt oli protsess sama mis põhikaardi rasteriseerimise puhul, erinev oli *.grd faili eksportimine, kus tuli valida atribuut *MSLINK*, mis sisaldab mullakoodi, mille järgi programm hiljem neid määrama hakkas. Parandati ka vead: muldade andmebaasi täiendati puuduvate mullakoodide ja -šifritega, tõsteti ümber atribuute ja polügone ühelt kaardilehelt teisele ning muudeti nende väärtusi, et programm neid korrektselt loeks.

3.2.4. Maapinna kõrgusmudel SRTM

Bakalaureusetöös kasutati SRTM (*Shuttle Radar Topography Mission*) Balti kõrgusmudelit, mis katab nii Eesti kui ka selle naaberriikide alasid. Ebavajalik info eemaldati – lõigati välja Eesti osa ja hiljem bakalaureusetöös vaja olevate põhikaardilehtede mõõtmete ja koordinaatidega lehed.

Kõrgusandmete kihi piksli algne küljepikkus oli 70 m, mistõttu ei saanud kogu Eestit hõlmavast kihist Eesti põhikaardi lehtede suurusi tükke välja lõigata. Probleemi lahendamiseks sobildati programmi *Idrisi32* abil nurgakoordinaate nii, et piksli küljepikkuseks sai 100 m, mis võimaldab lõikamist 10×10 km lehtedeks (s.t. 100 m jagub 10 000 m-ga). Koolutamisel kasutati suurt, kogu Eestit ning vähesel määral ka naaberalasid hõlmavat kihti, mis tagas selle, et suuremad moonutused esinesid servaaladel ning keskel, kus asusid uuritavad alad, olid need võimalikult väikesed.

3.2.5. Ortofotod ja asendustunnuste kihid

Uurimuses kasutati 2005. a ortofotosid, kuna need olid olemas mõlema uuritava ala kohta. Vajalikule kujule teisendatud ning kihtideks lahutatud ortofotod kasutatavate kaardilehtede kohta olid andmekihtide hoidlas olemas. Ortofotokihid olid: punane (R, *red*), roheline (G, *green*), sinine (B, *blue*), kollane (Y, *yellow*), värvitoon (H, *hue*), küllastus (S, *saturation*) ja heledus (L, *lightness*).

Asendustunnused on eelnevalt omaette andmekihtidena välja arvutatud kihid, mis muudavad prognoosikaartide genereerimise kiiremaks. Kaardi genereerimise protsessi kiirust mõjutavad mitmed näitajad, üheks olulisemaks on ortofototunnused, mille pikslisuurus on väga väike (1×1 meetrit), mis muudab arvutused väga aeglaseks. Seetõttu, kasutades programmi *Pidevstuudium*, arvutati valmis enamiku ortofototunnuste rasterfailid resolutsiooniga 10×10 m.

4. Pidevstuudiumi metoodika

Tehisõppesüsteem *Pidevstuudium* on loodud Tartu Ülikooli geograafia osakonnas järgmistel eesmärkidel (Remm, Linder, 2007; Remm, Remm, 2008):

- 1) lokaalsete ruumimustri indeksite arvutamiseks,
- 2) sarnasusele tuginevaks prognoosimiseks vajalike parimate tunnuste ja näidiste väljaotsimiseks ning nende kaalude optimeerimiseks tehisõppe teel,

3) nominaalsete ja pidevate nähtuste hindamiseks tabelisse või rasterkaardile.

Süsteem *Pidevstuudium* koosneb kolmest osast: 1) andmekihid; 2) vaatluste ja tehisõppe ning hinnangute andmebaas(id); 3) tehisõppetarkvara *Pidevstuudium*. On vajalik, et andmebaaside ülesehitus vastaks *Pidevstuudiumi* juhendis (Remm, Linder, 2008) kirjeldatule, kasutatavate andmete asukohad oleksid kindlaks määratud ja väljundkaustad loodud.

Pidevstuudiumit on geograafia osakonnas kasutatud mitmes uurimuses ja rakenduslikes töödes, milles on prognoositud taimkatet/maakatet ja kaitsealuste liikide esinemist (nt Linder *et al*, 2008; Remm, 2004, 2005, 2006; Remm *et al*, 2007), aga ka otseselt ruumiga mitteseotud nähtusi (Remm, Remm, 2008).

4.1. Ruumimustrit kirjeldavad tunnused

Bakalaureusetöös kasutatud 100 erinevat tunnust on toodud lisas 1. Tunnused jagunesid andmekihitüüpide vahel järgmiselt: põhikaardilt arvatati 8, kõrgusmodeli andmekihist 7, ortofotodelt 60 (neist asendustunnused 18), mullakaardilt 9 ning satelliidipiltidelt 16 tunnust.

Pärast seda, kui andmekihid olid ette valmistatud ja andmebaas korda seatud, hakati kihtidest programmiga *Pidevstuudium* argumenttunnuseid arvutama. Tunnuste arvutamisel kasutati sarnasuse nullkaugust. Sarnasuse nullkaugus reguleerib teineteisest teatud kaugusel asuvate vaatluste võimalust olla näidiseks üksteie prognoosimisel. Ruumilise autokorrelatsiooni tõttu on lähemal olevad vaatlused sarnased, see võib põhjustada aga prognoosimise täpsuses vigu – liiga kõrgeid hinnangu tõesuse näitajaid, sest lähestikku paiknevad vaatlused prognoosivad üksteist väga hästi (Mather, 1999; Remm, Linder, 2008). Nullkauguseks määrati 50 m. Tunnused arvutati kõigile õpetusvaatlustele. Mõnede tunnuste puhul piiras tunnuste arvutamist eelklassifikaatorikiht, milleks sedakorda oli eelnevalt rasteriseeritud vaatlusalade piiride kiht.

4.2. Tehisõpe

Kõiki olemasolevaid tunnuseid ning näidiseid pole mõtet hinnangutes kasutada, kuna need võivad üksteist dubleerida, olla aegunud või ebausaldusväärsused. Seetõttu viiakse läbi tehisõpe, mis on protsess, kus otsitakse vaatlusandmetele võimalikult hästi vastavaid hinnanguid andvate näidiste ja tunnuste kaalusid. Selle peamised ülesanded on (Remm, Linder, 2008):

- 1) leida iga uuritava nähtuse prognoosimiseks parimad näidised;
- 2) leida parim tunnuste komplekt uuritava nähtuse äratundmiseks.

Iseenesest on tehisõpe üsna aeganõudev protsess, mida mõjutab näiteks kasutatavate näidiste (vaatluste) hulk ning iteratsioonide arv. Näiteks võttis 2,7 GHz Intel Pentium Core 2 Duo protsessori ja 2 GB 400 MHz DDR2 RAM-i ning protsessi prioriteedi *Normal* (vaikimisi *Low*) juures 449 vaatlusega valimi õppimine 1000 iteratsiooni puhul aega 3 päeva ja 1 tund, sellest aga umbes viis korda väiksema (93) valimi puhul kõigest 2,5 tundi.

Enne tehisõpet korrastati tehisõppe andmebaas vastavalt *Pidevstuudiumi* juhendile (Remm, Linder, 2008). Kõigi valimite puhul määrati tunnuste kaalumiste arvaks 20, vaatluste kaalumiste arvaks üks ning iteratsioonide arvaks 1000 korda.

4.3. Hinnangute arvutamine

4.3.1. Hinnangud andmebaasi tabelisse

Antud juhul hinnati 11 taimkatteklassi (tabel 2). Iga valimi jaoks tehisõppe käigus leitud parimat näidiste ja tunnuste komplekti kasutas *Pidevstuudium* prognooside ja ristprognooside arvutamisel. Kõigepealt prognoositi Karula vaatlusalade funktsioontunnuseid vaid Karula vaatlusi sisaldavate valimitega (nn *siseprognoosid*), siis vaid Otepää vaatlusi sisaldavate valimitega (nn *ristprognoosid*) ning seejärel segavalimeid sisaldavate vaatlustega (*osalised siseprognoosid*, *osalised ristprognoosid*). Seejärel tehti prognoosid samade valimitega Otepää uurimisala kohta ning lõpuks mõlema ala kohta kokku.

Andmebaas, mida prognoosil kasutatakse, peab vastama *Pidevstuudiumi* juhendis (Remm, Linder, 2008) kirjeldatud ülesehitusele ja läbitud peavad olema eelnevad etapid. Andmebaasi tabelisse salvestatakse nii prognoositava klassi number kui ka hindamisel kasutatud näidiste keskmine sarnasus vaatlusega.

4.3.2. Hinnangud binaarfaili

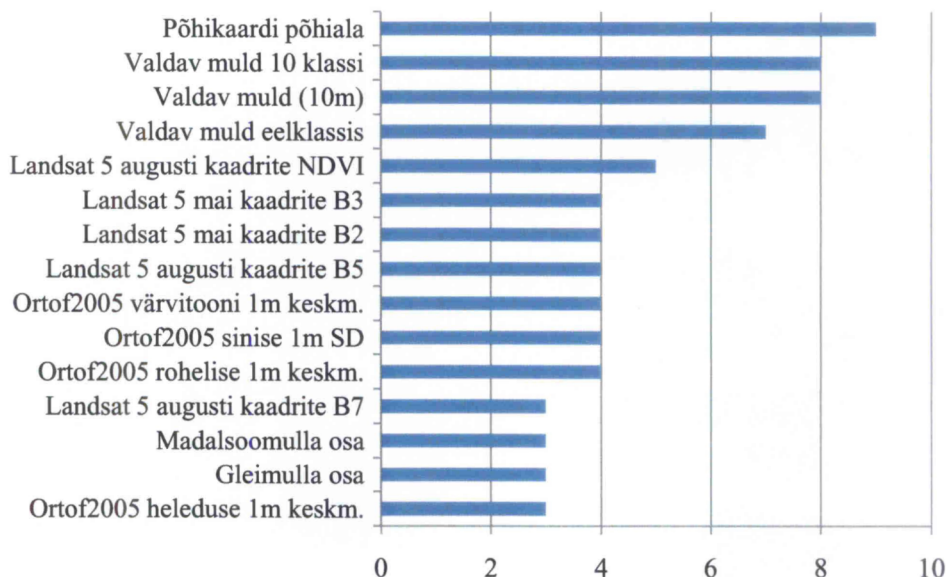
Väljundfailiks olev binaar-raster on Idrisi *rst*-vormingus rasterkaart. Prognoosikaart valmib sammudena ja on üsna aeganõudev; ajakulu annab vähendada asendustunnuste abil (vt. ptk. 3.5). Kaardi valmimise protsessi on täpsemalt kirjeldatud *Pidevstuudiumi* juhendis (Remm, Linder, 2008).

Prognoosikaardi näide ning sellega kaasnev hinnangute otsusekindluse (näidistega sarnasuse tase otsuse langetamisel) kaardi näide genereeriti kaardilehtede 4494 ja 5404 kohta, kasutades interpoleerimiskihina põhikaardi 10 m resolutsiooniga samade lehtede kihte. Prognoosikaardilt arvutati ka saadud taimkatteklasside levik, milleks maskiti eelnevalt, kasutades programmi *Idrisi32* ning põhikaarti, välja teed, järved, hoonestus ja muud alad, mis ei käinud taimkatteklassifikatsiooni alla.

5. Tulemused

5.1. Tehisõppel välja valitud tunnustest

Tehisõppe tulemusel valiti välja parimad andmekihtide tunnuste komplektid, mille põhjal teha prognoose, sellist parimate tunnuste välja sorteerimist, mille põhjal hinnanguid anda, mainib ka Pekkarinen (2004). Tehisõppe käigus valitakse välja uuritavate klasside prognoosimiseks vajalikumad tunnused. Kõige sagedamini parimas komplektis olnud tunnused üle kõigi valimite on toodud joonisel 5.



Joonis 5. 15 enim kasutatud tunnust parimates komplektides.

5.2. Hinnangud tabelisse

Iga valimi prognoosimisel saadud hinnangute ning välitööl määratud klasside kokkulangevused on toodud tabelis 3. Heleoranžid lahtrid tähistavad siseprognoose – ühe piirkonna valimiga prognoositi teist sama piirkonna valimit või tervet piirkonda, kust vaatlused pärit olid. Keskmiseks siseprognooside kokkulangevuseks oli 0.62, ehk 62% vaatlustest langesid prognoosiga samasse klassi milleks nad välivaatluste tõlgendamisega määrati. Parim tulemus saavutati Karula valimiga *kar2* selle sama valimi prognoosimisel (0.98). Halvim tulemus (0.39) saadi Otepää ja Karula segavalimiga *ote&kar1* teise segavalimi – *ote&kar2* – prognoosimisel.

Helerohelised kastid (tabel 3) tähistavad ristprognoose (Karula valimiga Otepää hinnangud või vastupidi) mille tulemused olid märgatavalt halvemad kui enese prognoosimisel. Keskmise täpsus kõikide ristprognooside kohta oli 0.36, mis jäi isegi madalamaks kui siseprognooside kõige viletsam täpsus. Parimaks oli 0.43 ja halvimaks 0.28.

Helehallid lahtrid (tabel 3) tähendavad segavalimiga ühe piirkonna valimite prognoosimist. Nende prognoositäpsused olid tunduvalt paremad kui ristprognoosidel, jäädes siiski alla siseprognooside kokkulangevustele. Minimaalne vastavus oli nende hulgas 0.42, keskmine 0.51 ja maksimaalne 0.63.

Helesinised lahtrid (tabel 3) on prognoositäpsused ühe piirkonna valimi poolt segavalimite vaatluste prognoosimisel. Need tulemused olid natuke paremad kui vastupidisel protsessil (segavalimiga ühe ala vaatluste prognoosimine), jäädes siiski alla enda ala klasside hindamise täpsusele. Minimaalne, keskmine ja maksimaalne tulemus olid vastavalt 0.41, 0.52 ja 0.68.

Tabel 3. Karula ja Otepää uurimisalade prognoosid ja ristprognoosid. *kar1*, *kar2* – 93-vaatluselised Karula juhuvalimid, *ote1*, *ote2* – 93-vaatluselised Otepää juhuvalimid, *ote&kar1*, *ote&kar2* – 93-vaatluselised Otepää ja Karula segavalimid; *kogukarula93*, *koguotepaa93*, *koikvaatlused93* – vastavalt 89, 85 ja 100 vaatlust sisaldavad (*Pidevstuudiumis* määratud kasutama umbes 93-vaatluselisi valimeid) Karula, Otepää ja kõigi vaatluste hulgast võetud prognoosivalimid; *kogukarula*, *koguotepaa*, *koikvaatlused* – kõiki vastava uurimisala vaatlusi sisaldavad prognoositavad valimid. Hele-rohelisega on tähistatud ristprognoosid (Karula prognoosid Otepää valimitega ja vastupidi), heleoranž – siseprognoosid (sama ala, kust prognoosivalim pärit, prognoosid), helehall – osalised ristprognoosid, s.t. väliste lisavaatlustega prognoosid (Otepää/Karula prognoosid Otepää ja Karula segavalimitega), helesinine – osalised siseprognoosid (mitme ala prognoos ühe ala valimiga).

		Prognoositav valim								
		kar1	kar2	ote1	ote2	ote&kar1	ote&kar2	kogu-karula	kogu-otepaa	koik-vaatlused
Prognoosi aluseks olev valim	valimi maht	Valimi maht								
		93	93	93	93	93	93	243	206	449
kar1	93	0.75	0.54	0.43	0.38	0.68	0.46	0.62	0.41	0.52
kar2	93	0.58	0.98	0.28	0.29	0.62	0.51	0.72	0.30	0.53
ote1	93	0.40	0.42	0.85	0.51	0.57	0.55	0.41	0.65	0.52
ote2	93	0.39	0.37	0.52	0.83	0.56	0.49	0.37	0.65	0.50
ote&kar1	93	0.51	0.51	0.55	0.53	0.74	0.39	0.46	0.52	0.49
ote&kar2	93	0.51	0.53	0.47	0.51	0.48	0.78	0.51	0.48	0.49
kogukarula93	89	0.60	0.62	0.29	0.33	0.54	0.41	0.62	0.29	0.47
koguotepaa93	85	0.32	0.40	0.58	0.65	0.47	0.45	0.36	0.60	0.47
koikvaatlused93	100	0.55	0.63	0.44	0.44	0.58	0.54	0.59	0.42	0.51

Hinnangute kokkulangevusi võrreldi omavahel kasutades programmi *Statistica 8 Mann-Whitney U testi*. Ristprognooside täpsuste võrdlemine teist tüüpi prognoosidega näitas, et saavutatud keskmised olid usaldusväärselt erinevad ($p < 0.05$ kõigil juhtudel). Ristprognooside (Karula põhjal Otepää ja Otepää põhjal Karula hinnangud) keskmisi võrreldes leiti, et hinnangute täpsus usaldusväärselt ei erine ($p = 0.1$).

Väljastpoolt prognoositavat maastikurajooni pärinevate õpetusalade lisamine prognoosivalimisse ja/või lisaks piirkonnale, kust õpetusvaatlused pärit, kaugema piirkonna prognoosimine, s.o. osalised rist- ja osalised siseprognoosid, andsid keskmisi, kuid

ristprognoosidest siiski tunduvalt paremaid tulemusi (kokkulangevuste keskmised vastavalt 0.51 ja 0.52). Nende kahe grupi keskmised omavahel usaldusväärselt ei erinenud ($p=0.76$).

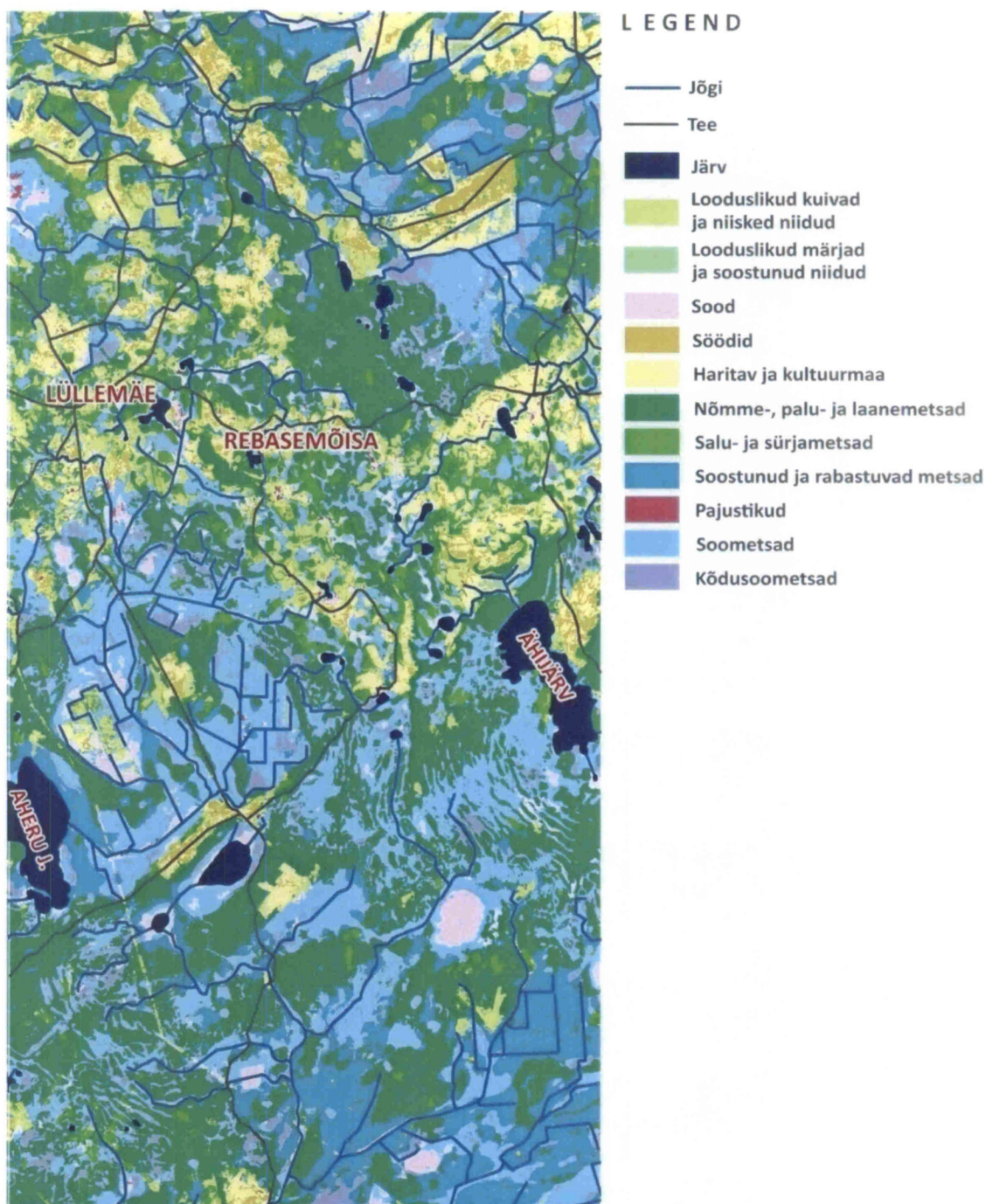
5.3. Prognoosikaart

Esialgse näitliku prognoosikaardi genereerimiseks valiti tehisõppe läbinud valim, mis sisaldas vaatlusi nii Karulast kui Otepäält ning kus olid esindatud kõik funktsioontunnuse klassid.. Näidiskaardid Karula ala kaardilehtede 4494 ja 5404 kohta valmisid etappidena, mille viimaseks sammuks oli valmis prognoosikaart (joonised 6, 7).

Prognoosikaardilt arvatud taimkatteklasside pindalasid (tabel 4) uurides selgub, et mõlemal kaardilehel on levinuimaks klassiks ülekaalukalt nõmme-, laane- ja palumetsad, millele järgnevad soometsad. Lõuna pool ehk kaardilehel 4494 on metsade ja seejuures soo- ja soostunud ning rabastunud metsade osakaal tunduvalt suurem kui põhja pool (kaardilehel 5404). Põhjapoolsel kaardilehel on aga tunduvalt rohkem haritavat ja kultuurmaad, mis on ka tegelikkuses teadaolevalt nii (Karula...).

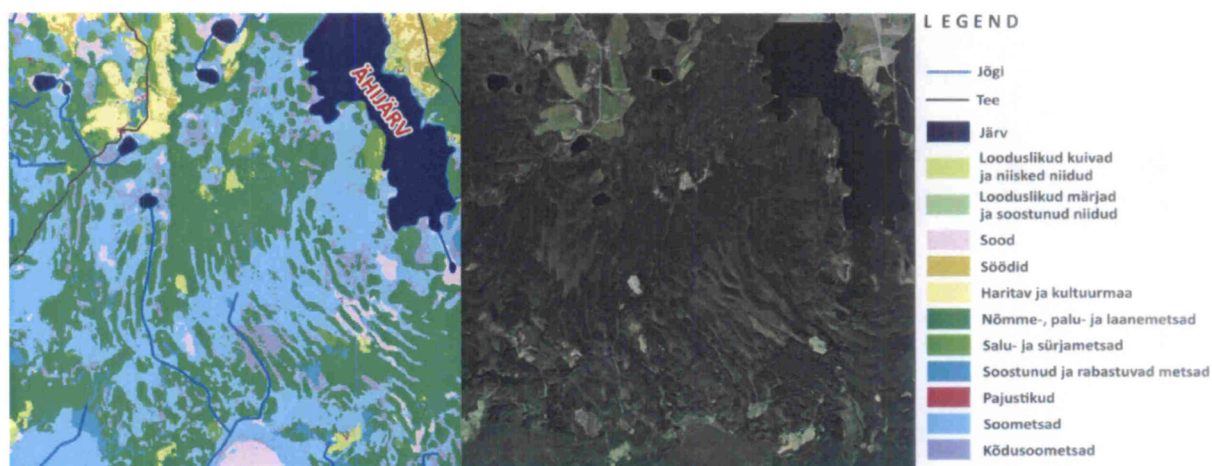
Tabel 4. Taimkatteklasside levik Karula kaardilehtedel 4494 ja 5404 esialgse prognoosikaardi järgi.

Klassi kood	Klassi nimetus	Kaardileht	Kaardileht
		4494	5404
		km ²	
1	Looduslikud kuivad ja niisked niidud	1.58	8.1
2	Looduslikud märke ja soostunud niidud	0.24	1.89
3	Sood	3.63	4.2
4	Söödid	1.02	5.68
5	Haritav ja kultuurmaa	1.66	12.15
6	Nõmme-, laane- ja palumetsad	39.08	29
7	Sürja- ja salumetsad	2.23	8.69
8	Soostunud ja rabastuvad metsad	13.49	6.4
9	Pajustikud	0.14	0.68
10	Soometsad	23.61	14.32
11	Kõdusoometsad	7.51	6.34



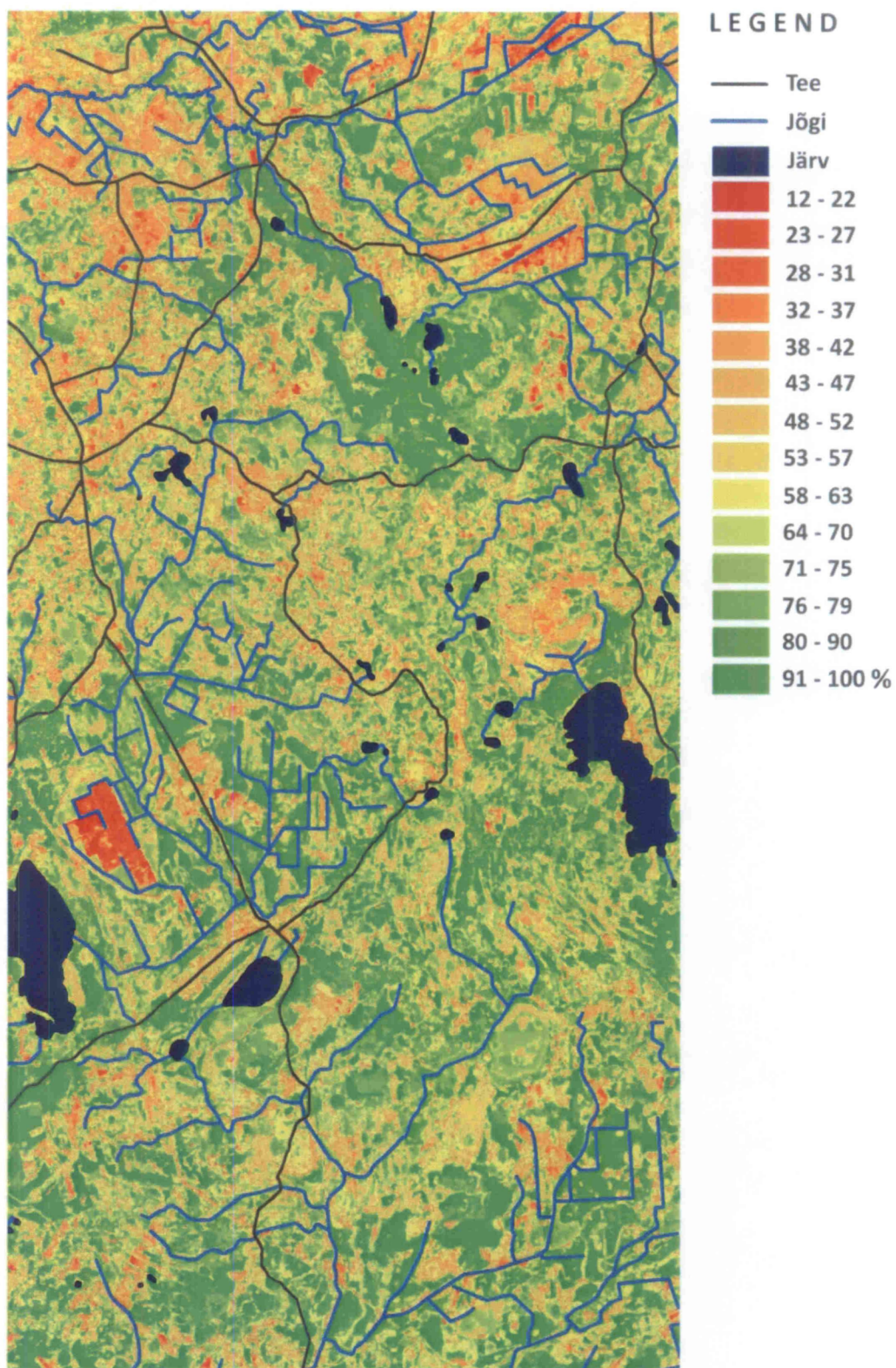
Joonis 6. Eesti Põhikaardi lehtede (1:10 000) 4494 ja 5404 11 taimkatteklassi prognoosikaart. Põhikaardilt on lisatud jõed, teed ja järved.

See, milline on maastik ortofotol ja millisena näeb ta välja peale prognoosikaardi genereerimist, võib näha allolevalt jooniselt 7.



Joonis 7. Ähijärve ümbruse taimkatte prognoosikaart (vasakul) ning sama ala 2005. a ortofoto (Eesti 1:10 000 põhikaardi kaardileht 44944). Põhikaardilt on lisatud jõed, järved.

Pidevstuudium loob ka ühe osana prognoosikaardi genereerimisel hinnangute otsusekindluse kaardi, kus väärtused on skaalal 0...100%. Kaartide loomisel kasutatavatel lehtedel 4494 ja 5404 oli minimaalne otsusekindlus vastavalt 12 ja 13% ning otsusekindluse graafiline kujutus on näha joonisel 8.



Joonis 8. Karula kaardilehtede 4494 ja 5404 hinnangute otsusekindluse kaart. Põhikaardilt on lisatud teed, jõed ja järved.

6. Arutelu

Kokku kasutati ettevalmistatud andmekihtide 100 tunnusest 61, neist 15 rohkem kui kahel korral ning ühte tunnust koguni 9 korral (joonis 5). Esiviisikus kasutatavuselt olid 3 tunnust mullakaardilt, mis viitab taimkatte ja mulla vahelisele tugevale seosele ning üks määrab teise, muld peamiselt maakatte tüübi (Yuan *et al*, 2006). Kõige enam kasutati aga Eesti põhikaardi põhialasid, mis määrasid suuresti, kus missugune maakasutus võib esineda.

Kõikide valimitega leitud parimad hinnangukomplektid võeti kasutusse prognooside tegemisel ja kaardi genereerimisel. Arvutati hinnangute kokkulangevused välitöödel määratud klassidega. Parimaks tulemuseks saadi 0.98 valimi *kar2* siseprognoosi puhul. Nii kõrget prognoositäpsust võib põhjustada tehiseõppe kasutamisel sageli esinev üleõppimine ehk ülesobitamine (*overfitting*), mis tähendab, et valim õpitakse arvuti poolt niivõrd selgeks, et samas valimis olevaid vaatlusi oskab ta prognoosida väga kõrge täpsusega, aga kui on vaja hinnata valimiväliseid kohti, esineb seal liiga palju segavat andmestikku (müra), mis muudab hinnangud ebatäpseteks (Dietterich, 1995; Tetko *et al*, 1995). Üleõppimisoht on seda suurem, mida väiksemad on õppevalimid (Dougherty, 2005). Et sellist olukorda vältida või sellest tulenevaid vigu vähendada, võib kasutada õppimise varasemat lõpetamist, ristkontrolli või tõenäosuslikkuse sidumist lisainfoga (Tetko *et al*, 1995). Kui valimiga *kar2*, millega saavutati juba mainitud siseprognoosi hinnangute kõrgeim vastavus, teist Karulast pärit sama mahuga ning samades kihtides sama arvu vaatlusi omavat juhuvalimit prognoositi, tuli täpsus vaid veidi üle poole. Siiski, kasutades seda valimit kõigi Karula vaatluste prognoosimiseks, saavutati samuti suhteliselt kõrge vastavus, mis võib sedakorda viidata asjaolule, et valimis õpiti selgeks teatud vaid Karulale omased tunnusjooned, mida väljaspool seda ala ei olegi võimalik prognoosida.

Ka Tokola (2000) on leidnud, et kasutades andmestikku väljastpoolt prognoositavat ala, suureneb hinnangu ebatäpsus vaatlustest kaugenedes, kui aga kasutada ka andmeid prognoositava ala seest, on saadud hinnangud märgatavalt paremad.

Ristprognoosiga saavutati halvemaid tulemusi kui siseprognooside või segavalimitega hinnangute puhul. *Mann-Whitney U test* näitas ristprognoosidega saavutatud hinnangute vastavuste võrdlemisel teiste prognoositüüpidega usaldusväärset erinevust.

Kui prognoosivalim koosnes nii prognoositava ala kui ka sellest väljaspool asuvatest vaatlustest, saavutati paremaid prognooside kokkulangevusi (0.51 ja 0.52) kui ristprognoosi puhul. Need keskmised usaldusväärselt ei erinenud ($p=0.76$). Sellest, et mingi ühe teatud piirkonna prognoos selle piirkonna ja väljaspool paikneva ala vaatluste segavalimiga andis usaldusväärselt paremaid tulemusi kui sama piirkonna ristprognoos vaid teise ala valimiga, võib oletada, et segavalimid sisaldavad lisaks teatud hulka heterogeensust, mis prognoositaval alal veel kirjeldamata.

Prognoosikaardil saadi halvima otsusekindluseks 12 – 13 %, mis viitab sellele, et madalaim näidistega sarnasuse tase otsuse langetamisel oli 12 -13 protsenti. Kõrgeimaks aga saavutati 100%, kus prognoositav piksel sarnanes näidisele täielikult. Keskmisteks olid 63 % ja 70 % vastavalt 5404-I ja 4494-I. See näitab, et enamasti oli otsusekindlus üle poole ja hinnang oli tõenäolisemalt õige, seda võib ka näha jooniselt 9, kus esineb rohkem kollakaid ja rohelisi toone, mis viitavad kõrgemale otsusekindlusele.

Prognoosikaardi järgi oli Karula kaardilehel 5404 haritav maa üsna sageli esinev klass. Kuigi hõredamini asustatud kaardilehe 4494 alaga võrreldes, kus on palju veekogusid ja metsa, on põhjapoolsel lehel ka teadaolevalt kultuurilade osakaal suurem. Siiski võivad nii mõnedki sellesse klassi hinnatud kohad sel kaardilehel aga kuuluda hoopis samuti üsna levinud looduslike kuivemate niitude või söötide hulka, kuna praeguste (liiga väheste) andmete järgi on nimetatud klassid, mis juba looduses on raskesti eristatavad, arvuti jaoks raskesti äratuntavad. Pindalaliselt kõige väiksemaks hinnati pajustike esinemisala, mis on aga usutav, kuna puu ja põõsarindes pajuliikide enamusega kooslused esinevadki sageli väikesepindalaliste laikudena soistes ja märgades lohkudes.

7. Kokkuvõte

Ruumiliselt eemal paikneva piirkonna taimkatte prognoosimine mingi teise regiooni andmete põhjal ja ilma prognoositavas kohas välitöid teostamata oleks maa- ja/või taimkatte suuremõõtkavalisel kaardistamisel väga ressursisäästlik meetod. Selliste võimaluste uurimiseks võeti ette ka käesolev töö. Uurimisaladeks valiti kaks maastikuliselt mitmes mõttes sarnast piirkonda ning naaber-maastikurajooni – Otepää ja Karula kõrgustikud koos neid ümbritsevate aladega (neli Eesti põhikaardi 10×10 km lehte mõlema uurimisala kohta). Kuna need kaks ala on maastiku ja taimkatte poolest üsna sarnased, võinuks eeldada, et ühe andmete põhjal saab edukalt prognoosida teise taimkatet. Töös kasutati välivaatlusi koos ortofotode, satelliidipiltide, mullakaardi, põhikaardi ja kõrgusmudeliga, et siis näidistel tugineva järeldamine ning pidevõppe teel seda tõestada. Bakalaureusetööst selgus aga, et ristprognooside tulemused olid üsna kehvad ning ühe piirkonna vaatlusandmete põhjal teise ala taimkatte hindamine ei ole anna eriti suuri hinnangute ja vaatluste kokkulangevusi. Paremaid tulemusi saavutati alles siis, kui valimitesse kaasati ka prognoositavalt alalt pärit vaatlusandmeid. Siiski jäid ka need täpsustulemused kesisteks. Parima tulemuse sai siis, kui kasutati vaatlusandmeid, mis olid pärit ainult prognoositavalt alalt. Seega, kui tulevikus soovitakse luua kogu Karula Rahvsupargi või kõrgustiku suuremõõtkavaline taimkattekaart, peaks iga hinnatava klassi kohta olema näidiseid ka Karulast. Valimi suuruse ning selle muutmise mõju prognoositäpsusele pole siiski veel teada ning see nõuaks juba eraldi uurimust.

Tänuavaldused

Sooviksin tänada oma juhendajat, Madli Linderit, kes oli minuga väga kannatlik ja vastas igale küsimusele, mis kohati just kõige selgemad ja vajalikumad ei olnudki. Samuti väga suured tänud proffessor Kalle Remmile, tänu kellele sai võimalikuks *Pidevstuudiumi* kasutamine ja kes andis head nõu andmetöötluse poole pealt.

Veelkord, suured tänud inimestele, kes osalesid vaatlusandmete kogumisel.

Eraldi tänaksin kannatliku suhtumise eest oma vanemaid, kes mind viimasel ajal just väga tihti ei näinud ning ka oma toakaaslast, kes pidi taluma mu hilisõhtuseid hiire ja klaveri piinamisi.

Lisaks tänan veel Kiira Mõisjat, Tõnu Ojat, Viktorit, DonQ-d ja kõiki oma sõpru ja tuttavaid, kes sellel pingelisel ajal mind toetasid ja abistasid.



Summary

Using cross-estimations in vegetation recognition in Otepää and Karula research areas

The use of remote sensing data is getting more and more important in land cover and vegetation recognition reducing the need for fieldwork in extensive areas if effective automatic interpretation and recognition methods have been developed.

This study aimed to find out how effective would be the predictions of the vegetation in one area using samples from another (the so-called cross-estimations). The study was carried out in Karula and Otepää uplands in South-East of Estonia. If proven to produce acceptable results, it would limit the costs of time and resources needed for vegetation mapping.

Field observations, Estonian 1: 10 000 digital base map, 1: 10 000 digital soil map, orthophotos, Landsat 5 TM satellite images and digital elevation model (SRTM) were used in the software system *Constud* (*Pidevstuudium*) for calculation of map and image pattern indices, for machine learning and for prediction of eleven vegetation classes.

A total of 449 field observations was used. Stratified random sample sets were created – two mutually exclusive sets for both area consisting of 93 samples and two exclusive sets combining random samples from both areas. In addition, the three sets consisting of about 93 observations selected out by *Constud*, were used.

The cross-estimations of vegetation classes did not give results as good as hoped (mean accuracy 0.36). Adding observations from the area under estimation raised the accuracy (0.51), about the same result (0.52) was achieved when both areas with the sample from one region were estimated. Using the samples acquired only from the area to be estimated, gave the best results (mean accuracy 0.62). The mean accuracies were statistically significantly different in the cases of cross-estimations compared to the other estimation types and when comparing the self-predictions to the other types ($p < 0.05$, Mann-Whitney U test).

Also, the predictive vegetation map and the map showing the certainty of estimations the class of a pixel was generated with *Constud*.

Viidatud kirjandus

- Aaviksoo, K. 1995.** Vegetation of Endla Nature Reserve classified on the basis of Landsat TM data. – Aaviksoo, K., Kull, K., Paal, J., Trass, H. (toim.). *Consortium Masingii*. Tartu Ülikool, Tartu, 27-36.
- Aaviksoo, K., Kadarik, H., Möllits, K. 1997.** Endla looduskaitseala taimkatte kaart. *Eesti Loodus* 10, 400-402.
- Aaviksoo, K., Paal, J., Dišlis, T. 2000.** Mapping of wetland habitat diversity using satellite data and GIS: an example from the Alam-Pedja Nature Reserve, Estonia. *Proceedings of the Estonian Academy of Sciences Biology Ecology* 49(2), 177-193.
- Aastaraamat Mets 2006.** Keskkonnaministeerium, metsakaitse- ja metsauuenduskeskus. Tartu 2007. 189 lk.
- Arold, I. 2005.** Eesti Maastikud. Tartu Ülikooli Kirjastus. Tartu. 453 lk.
- Bartolucci, A.A., Bae, S.J., Singh, K.P. 2004.** Sample Size Requirements for Stratified Random Sampling of Agricultural Run Off pollutants in Pond Water with Cost Considerations Using a Bayesian Methodology. *Third World Conference on Agricultural and Environmental Statistical Application*.
- Dietterich, T. 1995.** Overfitting and Undercomputing in Machine Learning. *ACM Computing Surveys*, Vol 27, No 3, September 1995.
- Eesti põhikaardi 1: 10 000 digitaalkaardistuse juhend. 2002.** Kinnitatud Maa-ameti peadirektori käskkirjaga nr. 13, 27. veebruar 2002. a. Maa-ameti kartograafiabüroo.
- Hang, E. 1979.** Otepää kõrgustiku kõrgussuhted: Eesti NSV saarkõrgustike ja järvenõgude kujunemine. Tallinn, 29-41.
- Idrisi32 Help. 2002.** Clark Labs, Clark University.

- Kilpeläinen, P. Tokola, T. 1999.** Gain to be achieved from stand delineation in LANDSAT TM image-based estimates of stand volume. *Forest Ecology and Management* 124, 105-111
- Kimes, D.S., Holben, B.N., Hickeson, J.E., McKee, W.A. 1996.** Extracting forest age in a Pacific North West from Thematic Mapper and topographic data. *Remote Sensing of the Environment* 56, 133-140.
- Kuusk, A., Lang, M., and Nilson, T. 2004.** Simulation of the reflectance of ground vegetation in sub-boreal forests. *Agricultural and Forest Meteorology* 126 (1-2), 33-46.
- Laasimer, L. 1965.** Eesti NSV taimkate. Valgus. Tallinn. 398 lk.
- Lillesand, T.M., Kiefer, R.W. 2000.** Remote Sensing and Image Interpretation. Fourth Edition. Wiley. 724 lk.
- Linder, M., Remm, K., Proosa, H. 2008.** The Application of the Concept of Indicative Neighbourhood on Landsat ETM+ Images and Orthophotos Using Circular and Annulus Kernels. Ruas, A.; Gold, C. (Toim.). Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling in Montpellier, France. Springer (*in press*).
- Lyons, A. 2002.** An Exploration of Methods for Analyzing Land Cover in Southern Zambia.
- Mather, P.M. 1999.** Computer Processing of Remotely-Sensed Images: An Introduction. Second Edition. Wiley. 292 lk.
- Meiner, A. 1999.** Eesti maakate. CORINE Land Cover projekti täitmine Eestis. Tallinn.
- Nilson, T., Kuusk, A., Lang, M., Lükk, T., Anniste, J., Peterson, U. 1999.** Satelliidipiltidel baseeruva metsakorralduse alused. – Pidev metsakorraldus. *EPMÜ Metsandusteaduskonna Toimetised* 32. Tartu, 81-87.
- Pekkarinen, A. 2002.** Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82 (2002) 349–359.

- Pekkarinen, A. 2004.** Image segmentation in multisource forest inventory. Finnish Forest Research Institute. Research papers 926.
- Raukas, A. 1995.** Eesti. Loodus. Valgus. Eesti Entsüklopeediakirjastus. Tallinn. 606 lk.
- Raukas, A., Rõuk, A.-M. 1995.** Pinnamood ja selle kujunemine. – Rmt.-s: Raukas, A. (koostaja). Eesti. Loodus. Valgus, Eesti Entsüklopeediakirjastus, Tallinn, 123-131.
- Remm, K. 2004.** Case-based predictions for species and habitat mapping. *Ecological Modelling* 177 (2004) 259–281.
- Remm, K. 2005.** Correlations between forest stand diversity and landscape pattern in Otepää Nature Park, Estonia. *Journal for Nature Conservation* 13 (2005) 137–145.
- Remm, K. 2006.** Liikide kasvukohad ja elupaigad hinnangulistel kaartidel. *Eesti Loodus*, 7, 36 - 39.
- Remm, K., Linder, M. 2007.** Prognoosisüsteemi *Pidevstuudium* tutvustus. *Geodeet* 34 (58) 2007.
- Remm, M., Remm, K. 2008.** Case-based estimation of the risk of enterobiasis. *Artificial Intelligence in Medicine*. doi:10.1016/j.artmed.2008.04.002
- Tetko, I.V., Livingstone, D.J., Luik, A.I. 1995.** Neural Network Studies. 1. Comparison of Overfitting and Overtraining. *Journal of Chemical Information and Computer Science* 35, 826-833.
- Tokola, T. 2000.** The Influence of Field Sample Data Location on Growing Stock Volume Estimation in Landsat TM-Based Forest Inventory in Eastern Finland. *Remote Sensing of Environment*. 74, 422–431.
- Yuan, J.-G., Fang, W., Fan, L., Chen, Y., Wang, D.-Q., Yang, Z.-Y. 2006.** Soil Formation and Vegetation Establishment on the Cliff Face of Abandoned Quarries in the Early Stages of Natural Colonization. *Restoration Ecology* 14 (3), 349–356.
- Zhen, G.Q., Zhang, S.T., Zhang, C.L., Fang, J.Y. 2004.** Automatic building detection from high resolution images based on multiple features.

Käsitirjad

Linder, M. 2006. Otepää looduspargi metsatüüpide kaardistamine näidistele tugineva pidevõppe tarkvara abil. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis.

Matsalu Rahvuspargi ja Nigula Looduskaitseala satelliitseire 1986-2001. 2007. Riikliku keskkonnaseire programmi „Maastike kaugseire“ aruanne. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis ning Info- ja Tehnokeskuses.

Oviir, M. 2007. Eesti põhikaardi okas-, sega- ja lehtmetsa eristatavus kaugseireandmete ja mullakaardi järgi kasutades näidistele tuginevat järeldamist ja tehiseõpet. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis.

Palo, A. 1999. Aerofotopõhine biotoopide kaardistamise meetod ja selle rakendamine looduskaitsealistes planeeringutes. Magistritöö maastikuökoloogia ja keskkonnakaitse erialal. 99 lk + lisad. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis.

Proosa, H. 2007. Optimaalne kerneli ulatus maakatteüksuste eristamiseks digitaalsest rastrist. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis

Remm, K. Ruumiliste loodusandmete statistiline analüüs. Versioon 3. Märts 2008. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis.

Remm, K., Linder, M. 2008. Tehiseõppesüsteemi Pidevstuumium kirjeldus. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis.
(http://map.gg.bg.ut.ee/kalle_r/PIDEVSTUUDIUM).

Remm, K., Palo, A., Linder, M. 2007. Otepää looduspargi taimkatte kaardistamise aruanne. Käsitiri Tartu Ülikooli Ökoloogia ja Maateaduste Instituudis ning
http://map.gg.bg.ut.ee/~kalle_r/Taimestikukaardistus/Otekaardiaruanne2007.pdf.

Internetiviited

GeoTIFF FAQ. <http://geotiff.maptools.org/faq.html>, 27. aprill 2008.

Karula Rahvuspark. <http://www.karularahvuspark.ee/>, 21. mai 2008.

Keskkonnaministeerium, 2008. <http://www.envir.ee/52897>, 12. mai 2008.

Landsat 5 Satellite Back in Action, 2008.
<http://www.usgs.gov/newsroom/article.asp?ID=1881>, 11. mai 2008.

LANDSAT General Overview, 2006. <http://www.landsat.co.za/overview.htm>, 11. mai 2008.

Otepää looduspark, 2006. <http://www.otepaaloodus.ee/>, 19. mai 2008.

Paal, J. 1997. Eesti taimkatte kasvukohatüüpide klassifikatsioon.
<http://www.botany.ut.ee/jaanus.paal/etk.klassifikatsioon.pdf>, 15. mai 2008.

Lisad

Lisa 1. Bakalaureusetöös kasutatud argumenttunnused ning nende arvutamisel kasutatud indeksid.

Argumenttunnuse nimi	Indeksi nimi
Põhikaardi põhiala	mood
Uusima põhikaardi 100 klassi	klasside arv
Põhikaardil kaugus järvest	kaugus etteantud klassi piirist pikslites
Põhikaardil kaugus teedest	kaugus etteantud klassi piirist pikslites
Põhikaardi serva lähedus	klassipiiri kaugus pikslites
Uusima põhikaardi 100 klassi külgnevus	klassikülgnevuste osa
Nõlvakalle (5pikslit)	gradiendi tugevus
Suhtkõrgus 5piksli keskmisest	erinevus keskmisest
Suhtkõrgus 10piksli keskmisest	erinevus keskmisest
Ekspositsioon (5pikslit)	gradiendi suund
Kõrgus (1piksel)	keskmine
Kaugusega kaalutud mulla mood	kaugusega kaalutud mood
Valdav muld (10m)	mood
Valdav muld eelklassis	mood
Mulla klassi külgnevuste osa	klassikülgnevuste osa
Ortof2005 punase 1m keskm.	keskmine
Ortof2005 punase 1m autokorr.	kaugusega kaalutud Morani I
Ortof2005 punase 1m gradiendi tugevus	gradiendi tugevus
Ortof2005 punase 1m naabri erinevus	naaberpikslite erinevus
Ortof2005 punase 1m naabruse autokorr.	Morani I 8st naaberpikslit
Ortof2005 punase 1m ekstsess	ekstsessitegur
Ortof2005 punase 1m sujuvus	üleminekute sujuvus
Ortof2005 punase 1m SD	standardhälve
Ortof2005 punase 1m max	maksimum
Ortof2005 roheline 1m keskm.	keskmine
Ortof2005 roheline 1m autokorr.	kaugusega kaalutud Morani I
Ortof2005 roheline 1m gradiendi tugevus	gradiendi tugevus
Ortof2005 roheline 1m naabri erinevus	naaberpikslite erinevus
Ortof2005 roheline 1m naabruse autokorr.	Morani I 8st naaberpikslit
Ortof2005 roheline 1m ekstsess	ekstsessitegur
Ortof2005 roheline 1m sujuvus	üleminekute sujuvus
Ortof200 roheline 1m SD	standardhälve
Ortof2005 roheline 1m max	maksimum
Ortof2005 sinise 1m keskm.	keskmine
Ortof2005 sinise 1m autokorr.	kaugusega kaalutud Morani I
Ortof2005 sinise 1m gradiendi tugevus	gradiendi tugevus
Ortof2005 sinise 1m naabri erinevus	naaberpikslite erinevus
Ortof2005 sinise 1m naabruse autokorr	Morani I 8st naaberpikslit
Ortof2005 sinise 1m ekstsess	ekstsessitegur

Ortof2005 sinise 1m sujuvus	üleminekute sujuvus
Ortof2005 sinise 1m SD	standardhälve
Ortof2005 sinise 1m max	maksimum
Ortof2005 kollase 1m keskm.	keskmine
Ortof2005 kollase 1m SD	standardhälve
Ortof2005värvitooni 1m keskm.	keskmine
Ortof2005värvitooni 1m SD	standardhälve
Ortof2005 küllastuse 1m keskm.	keskmine
Ortof2005 küllastuse 1m SD	standardhälve
Ortof2005 heleduse 1m keskm.	keskmine
Ortof2005 heleduse 1m autokorr.	kaugusega kaalutud Morani I
Ortof2005 heleduse 1m gradiendi tugevus	gradiendi tugevus
Ortof2005 heleduse 1m naabri erinevus	naaberpikslite erinevus
Ortof2005 heleduse 1m naabruse autokorr.	Morani I 8st naaberpikslist
Ortof2005 heleduse 1m ekstsess	ekstsessitegur
Ortof2005 heleduse 1m sujuvus	üleminekute sujuvus
Ortof2005 heleduse 1m SD	standardhälve
Ortof2005 heleduse 1m max	maksimum
Karbonaatmulla osa	antud klassi sagedus
Gleimulla osa	antud klassi sagedus
Madalsoomulla osa	antud klassi sagedus
Valdav muld 10 klassi	mood
Landsat 5 augusti kaadrite B1	keskmine
Landsat 5 augusti kaadrite B2	keskmine
Landsat 5 augusti kaadrite B3	keskmine
Landsat 5 augusti kaadrite B4	keskmine
Landsat 5 augusti kaadrite B5	keskmine
Landsat 5 augusti kaadrite B6	keskmine
Landsat 5 augusti kaadrite B7	keskmine
Landsat 5 mai kaadrite B1	keskmine
Landsat 5 mai kaadrite B2	keskmine
Landsat 5 mai kaadrite B3	keskmine
Landsat 5 mai kaadrite B4	keskmine
Landsat 5 mai kaadrite B5	keskmine
Landsat 5 mai kaadrite B6	keskmine
Landsat 5 mai kaadrite B7	keskmine
Landsat 5 augusti kaadrite NDVI	keskmine
Landsat 5 mai kaadrite NDVI	keskmine
Põhikaardi uusim metsaliik	mood
Põhikaardil kaugus kraavist	keskmine
Ortof2005 asendust. punase 30m keskm.	keskmine
Ortof2005 asendust. punase 30m SD	keskmine
Ortof2005 asendust. sinise 30m keskm.	keskmine
Ortof2005 asendust. sinise 30m SD	keskmine
Ortof2005 asendust. värvitooni 30m keskm.	keskmine
Ortof2005 asendust. värvitooni 30m SD	keskmine

Ortof2005 asendust. heleduse 30m keskm.	keskmine
Ortof2005 asendust. heleduse 30m SD	keskmine
Ortof2005 asendust. rohelise 30m keskm.	keskmine
Ortof2005 asendust. rohelise 30m SD	keskmine
Ortof2005 asendust. kollase 30m keskm.	keskmine
Ortof2005 asendust. kollase 30m SD	keskmine
Ortof2005 asendust. küllastuse 30m keskm.	keskmine
Ortof2005 asendust. küllastuse 30m SD	keskmine
Ortof2005 asendust. heleduse 30m naabri erinevus	keskmine
Ortof2005 asendust. heleduse 30m max	keskmine
Ortof2005 asendust. heleduse 30m gradiendi tugevus	keskmine
Ortof2005 asendust. heleduse 30m ekstsess	keskmine
Raskusjõutekkelise&tehnomulla osa	antud klassi sagedus
Nõlvakalle (10pikslit)	gradiendi tugevus
Ekspositsioon (10pikslit)	gradiendi suund
